Avrupa Bilim ve Teknoloji Dergisi
Sayı 38, S. 392-397, Ağustos 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 38, pp. 392-397, August 2022

Research Article

Performance Comparison of YOLOv3, YOLOv4 and YOLOv5 Algorithms: A Case Study for Poultry Recognition

Oğuzhan Kıvrak[1*], M. Zahid Gürbüz[2]

[1*] Bandirma Onyedi Eylul University, Bandirma Vocational School, Computer Programming Program, Balikesir, Türkiye, (ORCID: 0000-0001-5541-6749), okivrak@bandirama.edu.tr

[2] Dogus University, Engineering Faculty, Istanbul, Türkiye (ORCID: 0000-0002-5125-6378), zgurbuz@dogus.edu.tr

(İlk Geliş Tarihi 01 Mayıs 2022 ve Kabul Tarihi 21 Ağustos 2022)

(**DOI:** 10.31590/ejosat.1111288)

ATIF/REFERENCE: Kıvrak, O. & Gürbüz, M. Z. (2022). Performance Comparison of YOLOv3, YOLOv4 and YOLOv5 Algorithms: A Case Study for Poultry Recognition. *Avrupa Bilim ve Teknoloji Dergisi*, (38), 392-397.

Öz

Bu çalışmanın amacı, görüntüleri sınıflandırmak için kullanılan popüler evrişim sinir ağı modellerinin arasındaki performans farklılıklarını bulmaktır. Bunun için, YOLO modelinin farklı versiyonları üzerinde bir vaka çalışması yürütüldü. Bu çalışma için yeni bir veri seti oluşturulmuştur. Oluşturulan veri setinde, 918 adet tavuk, horoz ve civciv görüntülerini içeren kümes hayvanı fotoğrafları bulunmaktadır. Veri kümesinin % 80'i eğitim % 20 test olarak ayrılmıştır. Eğitim ve test veri kümelerindeki kümes hayvanlarının görüntüleri manuel olarak etiketlendi. Eğitim veri kümelesindeki görüntüler YOLOv3-tiny, YOLOv3, YOLOv4-tiny, YOLOv4, YOLOv5s, ve YOLOv5x modelleri kullanılarak eğitim tamamlandı. Kümes hayvanı tespiti için YOLOv5' modeli ile elde edilen sonuçlar diğer popüler CNN mimarisi sahip olan YOLOv3 YOLOv4 modelleri ile karşılaştırılmıştır. Sonuç olarak YOLOv5x(XLarge Depth(derinlik)) modeli 0,5 IOU'da %99,5 ortalama hassasiyetle en yüksek doğruluk oranı bulunmuştur.

Anahtar Kelimeler: YOLO, Görüntü İşleme, Evrişimsel Sinir Ağları, Performans Karşılaştırma, Kümes Hayvanı Tanıma, Bilişim Sistemleri, Sistem Geliştirme

YOLOv3, YOLOv4 ve YOLOv5 Algoritmalarının Performans Karşılaştırması: Kümes Hayvan Tanıma İçin Bir Vaka Çalışması

Abstract

The aim of this study is to classify poultries using popular convolutional neural network models. The different YOLO models are experimented to find best YOLO models in terms of performance. For this purpose, a case study was conducted on different versions of the YOLO model. A new dataset has been described in this study. In the dataset, there are 918 photos containing chickens, cockerel, and chicks. The dataset split into %80 training set and %20 test set. The images of poultries in the training and test datasets were manually annotated and those in the training dataset were used to train the YOLOv3-tiny, YOLOv3, YOLOv4-tiny, YOLOv4, YOLOv5s, and YOLOv5x Models. The results of using YOLOv5 for poultry detection are compared with other popular CNN architectures, YOLOv3, YOLOv4 models. The results show that YOLOv5x (XLarge depth) model records the highest accuracy, resulting in a mean average precision at 0.5 IOU of %99.5.

Keywords: YOLO, Image Processing, Convolutional Neural Network, Performance Comparison, Poultry Recognition, Information Systems, System Development

[*] Sorumlu Yazar: okivrak@bandirma.edu.tr

1. Introduction

Nowadays, with the developing Graphics Processing Unit (GPU) technology, image processing and object detection applications can be adapted in many fields. Deep learning technologies have become more popular with the increase in computing and capacity of GPU processors. There are many studies such as the perception of people walking on the road (Ahmed and Jeon, 2021), the presence of weeds in the field (Şin and Kadıoğlu, 2019), the detection and counting of vehicles in traffic, and the detection and counting farm animal such as cow, goat and sheeps (Kıvrak et al., 2020).

While it is easy for humans to recognize objects that they know they have seen before, however it is more difficult for computers to distinguish objects. Supported by deep learning algorithms, the detection of objects can be done by computers with a high success rate. Detection and tracking of objects can be done by determining many attributes of the object (Tan et al., 2021, 160).

With the increase in labelled data, which is important in this field, deep learning algorithms have been used to give meaning to the data. A large amount of data is used and it provides the desired performance without the need to manually extract the feature of the image (Tian et al., 2019: 2).

Object detection, one of the sub-topics of image processing, has an important place in computer vision applications. Therefore, object detection algorithms are based on supervised learning and artificial neural networks which has great interest in deep learning (Jubayer et al., 2021; Mathew and Mahesh, 2021). Object detection is process of the locating of certain objects in an image or video (Mutludoğan,2020:17). In the process of detecting the object, the object must be found in the image and its position must be determined. Detection and recognition of objects in the image is one of the most researched problems in image processing due to reasons such as changes in pose of objects, complexity and class diversity (Şimşek et al., 2019:634).

One of the models used in deep learning is Convolutional Neural Networks (CNN). CNN is a special model of a multilayer artificial neural network inspired by biological processes. This model, which is designed to recognize patterns from the pixels of the image, is a feed-forward artificial neural network that combines feature extraction and classification (Dandıl et al., 2019, 181). Today, CNN is the most efficient and widely used for object detection (Estaban et al., 2021).

In this study, YOLOv3, YOLOv4 and YOLOv5 models and its sub models, which are deep learning models, were compared to determine the most suitable algorithm for detecting poultry. The rest of the work is organised as follows. In the second part, materials and methods are explained in the context of dataset, cnn models, evaluation metrics, software and hardware used. In the third part, the results obtained from the experiments are stated and give insights about them. In the last part, the result of the study are discussed and offer some suggestions for further research.

2. Material and Mehod

Materials and methods are discussed in this section. It describes the experiments in terms of the dataset, software, hardware, and evaluation metrics.

2.1. Dataset

A custom dataset was created by collecting chickens, cockerel, and chicks via google search tool. The dataset contains 918 images. An image can contain more than one poultry.

The poultry in the images in the dataset were labelled with the help of the LabelMe annotation tool (Kentaro,2016). Rectangular bounding box technique was used as the labelling method. Because, YOLO algorithms work with such rectangular labelled data. Labelling was completed by giving the class names of the relevant poultry name each drawn rectangular bounding box. An instance of image labelling with the labelme tool is shown in the Figure 1. The names of these classes are “chicken”, “cockerel”, “chick”. After the completion of the tagging process, all tagged images and text files containing the coordinates of the bounding boxes were collected in a folder. The number of labeled class of each poultry are follows: 514 chicks, 652 chickens, and 276 cockerels.

Figure – 1 An example of labelled data within the dataset

2.2. Convolutional Neural Network Architectures

Although there are many algorithms and technologies used in object detection, convolutional neural networks are the most efficient and common tool used in object detection (Estaban et al., 2021). YOLO series models are single-step target detection models based on CNN. The YOLO model differs from two-stage target models such as the Faster R-CNN algorithm by transforming the object detection problem into a regression problem. YOLO models use neural network to predict the coordinates, probability value and class of the bounding box of the object in the image (Chen et al.; 2021:5-6). YOLO models can be applied real - time object detection applications because of good performance in terms of speed and accuracy (Jintasuttisak et al.,2022:2). In this study, YOLOv3, YOLOv4, and YOLOv5 versions were used.

i) YOLOv3

YOLOv3 is the improved version of the YOLO model (Redmon and Farhadi, 2018). YOLOv3 can perform localization and classification in real time only with the help of a neural network. This feature allows it to be trained with real-time inputs and to perform detection with a high probability (Iyer et al., 2021, 1156:1157). While YOLOv3 uses a similar structure to the Feature Pyramid Network to recognize objects, it also uses the Darknet53 network as a feature extractor (Kılıç et al.,2020:35).

i) YOLOv4

Yolov4 released by Bochkovskiy et al. in 2020. The training process with a single GPU that most modern scientific models use with a large mini-batch size, slows down and becomes heavier. Yolov4 tries to solve this slowdown problem with an object detector trained on a single GPU with a small mini batch size (Dewi et al.,2021:97229). It expands the data set with a new mosaic data augmentation method, and introduced a new the positioning loss function as C-IoU. This loss function optimised the direction of increasing overlapping areas and made the network more inclined. Thus, improved the accuracy (Yu and Zhang,2021:4). C-IoU loss function is defined as (1).

$$L_{CIOU} = 1 - IoU + \frac{\rho^2(B_{pre}, B_{act})}{c^2} + \alpha v \qquad (1)$$

IoU stands for intersection over union which is described below. B_pre and B_act represent predicted and actual bounding boxes respectively. ρ represents the euclidean distance between these bounding boxes. c signifies the diagonal length of the smallest enclosing box covering the two boxes. α signifies a positive trade-off parameter; ν is a parameter about measuring the consistency of aspect ratio (Kumar et al.,2021:5).

ii) YOLOv5

YOLOv5 is a single-stage detector and region-based object detection network, and the fifth version of the open source YOLO algorithm developed by a firm called Ultralytics. This model was developed with an improvement on the PyTorch library (Estaban et al., 2021; Ieamsaard et al., 2021:429). Yolov5 uses an architectural structure consisting of 3 main parts: spine, head and neck. In the ESA layer on the backbone, the features of the input image at different scales are extracted. The neck part creates a feature map by using the features it receives from the spine and carries it to the next part, the prediction layer. In the head part, localization and classification are made with the features taken from the neck part (Iyer et al., 2021: 1157; Murat, 2021: 41)

Figure 2 – YOLOv5 Architecture Model (Fang vd.,2021:5394)

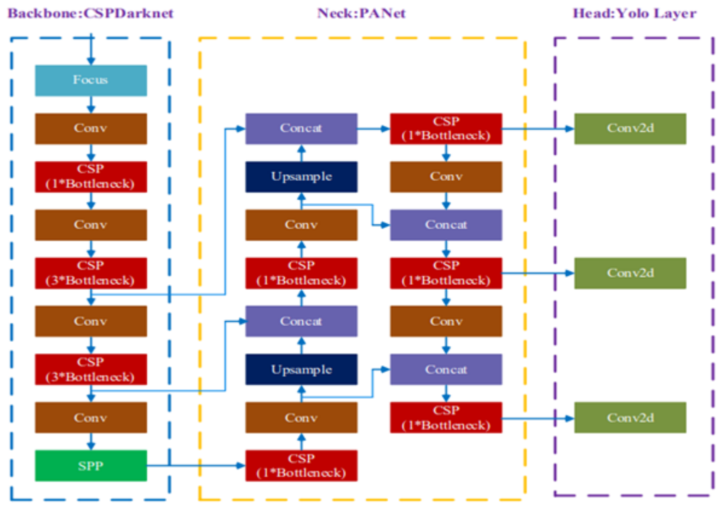


Table 1. List of softwares and hardwares

	Item	Specification
Software	Operating System	Ubuntu 18.04
	CUDA	11.1
	IDE	PyCharm 2021.1.3
	Programming Language	Python 3.8
Hardware	Processor (CPU)	Intel Core i9-10920X CPU @3.50 GHz 4x24
	Graphics Card (GPU)	2 x Nvidia GeForce RTX 3090 (24GB)
	Memory	128 GB

2.3. Software and Hardware

In the study, Python programming language was used as the primary software language and the PyCharm 2021.1.3 (Jetbrains, 2021) is used as integrated development environment (IDE). List of softwares and hardwares used in the experiments are listed at Table 1.

2.4. Evaluation Metrics

In this section, the evaluation metric used in the training of the models and the criteria used to determine the predictive validity of the models will be explained.

i) 4.1 Intersection over union (IoU)

Intersection Over Union (IoU) is the measure used to calculate the overlap of the area where the predicted bounding box and the actual bounding box intersect (Adrian, 2021). YOLO models use this metric to determine the overlap ratio of two bounding boxes. The domain of IoU is in the range of 1 and 0. 1 means that the predicted and actual bounding box is fully overlapped. The formula of IoU is described at (2) where R is the rectangle of bounding box, RA is actual bounding box and RP is predicted bounding box.

$$IoU = \frac{R_A \cap R_P}{R_A \cup R_P} \qquad (2)$$

In object detection, there are several bounding boxes for each object. IoU is calculated for each the predicted bounding boxes. Then the boxes are sorted according to this value. The boxes below the threshold is eliminated. If there are more then one box above the threshold, the box which has the maximum IoU value is selected. The threshold is defined as 0.5 or 0.95 in general. In this paper the threshold is defined as 0.5.

ii) Precision and Recall

Precision is the ratio of positive prediction value over all predictions. The formula of precision is (3). Recall is the ratio of positive prediction value over ground truth. The formula of recall is (4). In these formulas (3) and (4), TP refers to the predicted value that exists, FP refers to the predicted value that does not actually exist, FN refers to the non-predicted value that does not actually exist.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

iii) Mean average precision (mAP)

Mean average precision (mAP) refers to the mean average of the Average Precision (AP) values for all classes. Average Precision is the average of precisions of all predictions.

3. Experimental Results

In this section, different versions of YOLO models, including YOLOv3 (Redmon and Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), YOLOv5 (Cengil and Cinar, 2021), compared their performance on poultry detection. To compare the efficiency of each model, the dataset that contains poultry images, 6 YOLO models were run with the following parameters located at Table 2. Parameters were chosen as similar values for a fairer comparison.

The models were trained until they reach the max batch size or epoch size, and they were tested on identical train/test datasets. All experiments were done on the same PC with Intel Core i9-10920X CPU, 2 x Nvidia GeForce RTX 3090 GPUs, 128GB RAM. They were run separately to monitor the effect of the models under the same conditions.

Table 2. Parameters of YOLO Models

Parameters	Yolov3 Tiny	Yolov3	Yolov4 Tiny	Yolov4	Yolov5 s	Yolov5 xl
Number of iterations	max-batch: 6000	max-batch: 6000	max-batch: 6000	max-batch: 6000	epoch: 600	epoch: 600
Batch	64	64	64	64	64	64
Learning Rate	0.001	0.001	0.00261	0.0013	0.0013	0.0013
Momentum	0.9	0.9	0.9	0.949	0.949	0.949

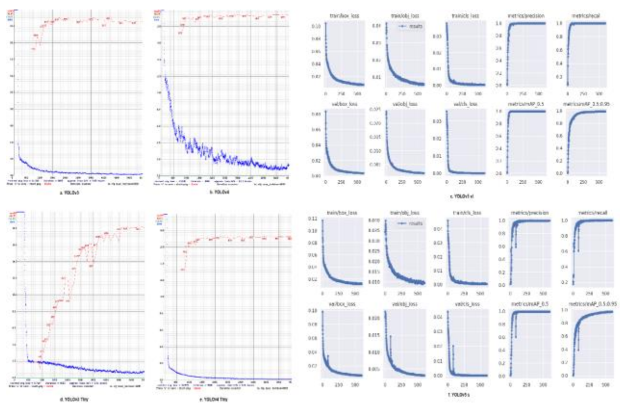
Table 3. The result of the experiments

Result	mAP@0.5 IoU	loss	Precision	Recall	Training Time (mins)
Yolov3 tiny	90.3	0.74	0.93	0.83	55
Yolov3	92.9	0.18	0.96	0.93	297
Yolov4 tiny	86.2	0.09	0.96	0.79	**48**
Yolov4	96.6	1.35	0.95	0.97	379
Yolov5 s	**99.5**	0.01	0.997	**0.99853**	108
Yolov5 xl	**99.5**	**0.007**	**1.00**	0.99783	279

Table 3 summarizes the results of all models. It can be seen that the best mAP value is %99.5 and it was obtained by both YOLOv5 models. The mAP values of other models are above %90 except YOLOv4 tiny model. In terms of training speed, the performance of YOLOv4 tiny model overcome the others. According to these results, although the training time of the YOLOv4 tiny model is the best, the mAP value is the worst. Tiny models have less convolutional network layers then the other versions. In terms of tiny models, YOLOv5s has double training time over others but it has overcome the YOLOv4 tiny and YOLOv3 tiny.

As seen in the figure 3, the loss curve of the YOLOv3 and YOLOv4 tiny model converges after 1800 steps with the value of below 1.00. The convergence speed decreases beyond this point. The experiment stop at 6000 iterations with the value of about 0.18 and 0.09. The loss curve of YOLOv4 model, converges to 1.35 at the end of the steps, however the mAP curve converges to about %96 at 1200 steps. The loss curve of the YOLOv3 tiny model converges after 5000 steps and as parallel to it mAP curve converges after this step. In the YOLOv5 models there are three types of loss shown in figure 3. The box loss indicates how well the predicted bounding box covers an object. Object loss is a measure of the probability of an object being present in a proposed region of interest. The cls loss represents how well the algorithm can predict the correct class. The all loss curves converges around epoch 500. Best weights can be obtained from the steps that converges.

Figure – 3 The results of all models



The examples in Figure 4 show that the YOLOv5s model can detect the poultries to a higher degree of certainty. In our dataset, even though it is hard to identify chicken and cockerel, the model detects them successfully. In some chick images, the model was not able to successfully identify all of them if there were many nested chicks. However, this is a situation that is difficult to detect with the human eye. It is also hard to identify in case objects are located far away from the camera.

Figure – 4 Images from the test dataset showing the performance for detecting poultry

4. Conclusions

In this study, YOLOv3, YOLOv4 and YOLOv5 models and its sub models are examined for performance comparison. For this purpose, a new custom dataset of poultries is described. The dataset contains 918 images that can contain more than one poultry in each image. The dataset has three classes as chickens, cockerel, and chicks. mAP metric is used to evaluate the results. The models were trained and tested on identical train/test datasets, and they used similar iteration numbers.

YOLOv5s and YOLOv5xl have the best performance in the context of mAP of 99.5% over others. mAP value of the YOLOv4 model is better than YOLOv3 and the tiny models. According to training time, although the YOLOv4 model has the best performance of 48 minutes, it has the worst mAP values. Thus, the YOLOv3 model can be used in terms of training time. Although the YOLOv4 model has the longest training time, it has a lower value than YOLOv5s and YOLOv5x by looking at the mAP metrics. Therefore, the YOLOv5x model or the YOLOv5s model should be preferred. For a good learning, it is expected that the loss values should be close to zero but not zero. The YOLOv5x model also has the best loss value. In the study of palm tree detection, although all YOLO models have similar training time, YOLOv5s model is minimum [26] (p. 8). In our case, there is a larger gap between the minimum (48 minutes) and maximum (379 minutes) training time.

In this study, yolov3, yolov4 and yolov5 were used. In future studies, the latest versions of Yolo, yolov6 and yolov7, can be used to compare performance differences. In addition, the models used in this study can be run with a different data set and the results can be compared.

Acknowledgments: The authors would like to thank the Dogus University Scientific Research Projects Coordination Department under the Grant No. 2019-20-D2-B07 for financial support and making the study available by providing hardware for our research.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

1. Ahmed, I., Jeon, G.(2021). A real-time person tracking system based on SiamMask network for intelligent video surveillance. J Real-Time Image Proc (2021). https://doi.org/10.1007/s11554-021-01144-5.
2. Şin, B. Kadıoğlu, İ. (2019). İnsansız Hava Aracı (İHA) ve Görüntü İşleme Teknikleri Kullanılarak Yabancı Ot Tespitinin Yapılması . Turkish Journal of Weed Science , 22 (2) , pp. 211-217 . Retrieved from https://dergipark.org.tr/en/pub/tjws/issue/51404/669501.
3. Kıvrak,O., Gürbüz M.Z., Güran, A. (2020), Çivi Yazısından Dijital Kodlamaya Sosyo -Ekonomi Çalışmaları, Ekin Yayınevi, pp.1-12.
4. Tan , F.; Yüksel ,A.; Aydemir ,E.; Ersoy , M. (2021). "Derin Öğrenme Teknikleri İle Nesne Tespiti Ve Takibi Üzerine Bir İnceleme", Avrupa Bilim ve Teknoloji Dergisi, no. 25, pp. 159-171, Jun. 2021, doi:10.31590/ejosat.878552.
5. Tian, Y., Yang, G. Wang, Z. Li,E.,Liang, Z. (2019). Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. Hindawi Journal of Sensors. https://doi.org/10.1155/2019/7630926.
6. Mathew, M.P., Mahesh, T.Y. (2021).Leaf-based disease detection in bell pepper plant using YOLO v5. SIViP (2021). https://doi.org/10.1007/s11760-021-02024-y.
7. Jubayer, M.F.; Soeb, M.J.A.; Paul, M.K.; Barua, P.; Kayshar, M.S.; Rahman, M.M.; Islam, M.A.(2021). Mold Detection on Food Surfaces Using YOLOv5. Preprints 2021, 2021050679 (doi: 10.20944/preprints202105.0679.v1).
8. Mutludoğan, K.2020. Derin Öğrenme Tabanlı Şeffaf Nesne Tanıma. Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi.
9. Şimşek, E., Özyer, B., Tümüklü Özyer, G., (2019). Fotokapan Görüntülerinde Yerel Öznitelikler ile Nesne Tespiti. Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi , cilt.9, sa.4, 633-644.
10. Dandıl, E , Turkan, M , Boğa, M , Çevik, K . (2019). Daha Hızlı Bölgesel-Evrişimsel Sinir Ağları ile Sığır Yüzlerinin Tanınması . Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi , BŞEÜ Fen Bilimleri Dergisi 6. Cilt - Prof. Dr. Fuat Sezgin Bilim Yılı Özel Sayısı , 177-189 . DOI: 10.35193/bseufbd.592099
11. Esteban, J.M., Loosdrecht, J.V., Aghaei, M. (2021). Obstacle Detection for BVLOS Drones. ArXiv, abs/2106.11098
12. Kentaro Wada. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme, 2016
13. Chen, J., Wang, Z., Wu, J., Hu, Q., Zhao, C., Tan, C., Teng, L., & Luo, T. (2021). An improved Yolov3 based on dual path network for cherry tomatoes detection. Journal of Food Process Engineering, 44(10), e13803. https://doi.org/10.1111/jfpe.13803
14. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

15. Iyer, Rakkshab & Bhensdadiya, Kevin & Ringe, Priyansh. (2021). Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for Real-Time Mask Detection. International Journal of Research in Engineering and Technology. 2395-0056.
16. Kılıç, B. , Baykal Kablan, E. , Doğan, H. , Ekinci, M. , Ercin, M. E. & Ersöz, Ş. (2020). Derin Konvolüsyonel Nesne Algılayıcı ile Plevral Efüzyon Sitopatolojisinde Otomatik Çekirdek Algılama . Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi , 13 (1) , 33-42 . Retrieved from https://dergipark.org.tr/en/pub/tbbmd/issue/53711/638455.
17. Fang, Y., Guo, X., Chen, K., Zhou, Z., and Ye, Q. (2021). "Accurate and automated detection of surface knots on sawn timbers using YOLO-V5 model," BioResources 16(3), 5390-5406.
18. Ieamsaard, J., Charoensook, S. N., Yammen, S., "Deep Learning-based Face Mask Detection Using YoloV5," 2021 9th International Electrical Engineering Congress (iEECON), 2021, pp. 428-431, doi: 10.1109/iEECON51072.2021.9440346.
19. Murat, S. (2021).İnsansız Hava Aracı Görüntülerinden Derin Öğrenme Yöntemleriyle Nesne Tanıma. Yüksek Lisans Tezi, Maltepe Üniversitesi.
20. Jetbrains. Pycharm. https://www.jetbrains.com/pycharm/, Erişim Zamanı: 23.08.2021.
21. Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. 2020. "Yolov4: optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934
22. Cengil, E., Çınar, A. (2021). Poisonous Mushroom Detection using YOLOV5. Turkish Journal of Science and Technology, 16(1), 119-127.
23. Dewi,C., Chen, R. -C., Liu, Y. -T. , Jiang, X. , Hartomo, K. D. , (2021),"Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN," in IEEE Access, vol. 9, pp. 97228-97242, 2021, doi: 10.1109/ACCESS.2021.3094201
24. Yu, J, Wei Z. (2021). Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4. Sensors 21, no. 9: 3263. https://doi.org/10.3390/s21093263
25. Kumar, A., Kalia, A., Sharma, A. et al. A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system. J Ambient Intell Human Comput (2021). https://doi.org/10.1007/s12652-021-03541-x
26. Jintasuttisak, T. Edirisinghe, E. El Battay, A. (2022). Deep neural network based date palm tree detection in drone imagery. Computers and Electronics in Agriculture. 192.10.1016/j.compag.2021.106560.